#InvestWatch – na straży inwestorów!

1. Wprowadzenie - opis organizacji, sytuacji i stanu aktualnego

Waszym zadaniem jest zaprojektowanie i zbudowanie kompleksowego systemu monitoringu treści w sieci (#InvestWatch), który pozwoli na automatyczną identyfikację, agregację i analizę materiałów promocyjnych oraz ofert dotyczących instrumentów i usług finansowych na polskim rynku. System powinien opierać się na uniwersalnym silniku web scrapingu oraz trzech wyspecjalizowanych modułach analitycznych, wspierających Urząd Komisji Nadzoru Finansowego w ochronie inwestorów.

---

2. Wyzwanie

Rozwiązanie powinno składać się z centralnego silnika zbierającego dane (z Internetu oraz mediów społecznościowych, m.in. Facebook, Instagram, TikTok) oraz następujących modułów analitycznych:

Moduł 1: Monitoring ekosystemu zewnętrznego (Finfluencerzy i podmioty trzecie)

- Zakres: Wyszukiwanie reklam zarejestrowanych w Polsce firm inwestycyjnych oraz oferowanych przez nie instrumentów i usług, które są zamieszczane przez podmioty trzecie lub osoby prywatne (w szczególności finfluencerów). Zakres obejmuje także polskojęzyczne reklamy zagranicznych firm inwestycyjnych działających w Polsce transgranicznie.
- Rezultat: Automatyczny raport/baza zestawiająca: dane firmy inwestycyjnej, dane podmiotu reklamującego, treść reklamy oraz jej dokładną lokalizację (URL) wraz z mechanizmem okresowej aktualizacji.

Moduł 2: Weryfikacja zgodności i informacji o ryzyku

- Zakres: Wyszukiwanie materiałów marketingowych publikowanych bezpośrednio przez zarejestrowane w Polsce firmy inwestycyjne (w tym polskie oddziały firm zagranicznych).
- Analiza (AI / Modele statystyczne): Wykorzystanie algorytmów do oceny, czy informacja o ryzyku towarzyszącym instrumentom lub usługom została zamieszczona w sposób dostateczny, proporcjonalny do informacji o zyskach oraz wyraźny (zgodnie z listą wymogów w załączniku).
- Rezultat: Narzędzie automatycznie flagujące materiały niespełniające kryteriów poprawności informacyjnej.

Moduł 3: Identyfikacja usług nieautoryzowanych (Shadow Advisory)

- Zakres: Wyszukiwanie opisów usług firm typu wealth management, investment consulting, family offices oraz innych podmiotów doradczych niebędących autoryzowanymi firmami inwestycyjnymi.
- Analiza: Wykrywanie fraz i treści mogących wskazywać na faktyczne świadczenie w Polsce usługi doradztwa inwestycyjnego bez wymaganych zezwoleń.
- Rezultat: Raport zestawiający dane podmiotu, treść oferty oraz lokalizację (URL) potencjalnego naruszenia.

---

3. Oczekiwany rezultat

Oczekiwanym rezultatem hackathonu jest działający prototyp (MVP) systemu, na który składa się:

1. Backend & Scraping Engine (.NET/Python/Node.js):

- Moduł pozyskiwania danych (crawler/scrapper):
  - Implementacja kolejki zadań w oparciu o dowolne narzędzie open source
  - Wykorzystanie bibliotek typu Playwright, Selenium, httpx, requests itp.
  - Implementacja kolektorów API lub parserów treści dla Social Mediów, przy zachowaniu rotacji User-Agent w nagłówkach i proxy
  - Ekstrakcja danych z HTML do formatu wyjściowego np. JSON.
- Warstwa składowania obiektów:
  - Baza metadanych (SQL)
  - Baza treści (SQL + np. Object Storage)
- Harmonogram:
  - CronJob jako mechanizm uruchamiania scrapperów lub system orchestratora.

2. Moduły Analityczne / AI (Python):

- Preprocessing:
  - Usuwanie tagów HTML, normalizacja tekstu, usuwanie duplikatów,
  - Reprezentacja wektorowa w postaci tokenów i embeddingów.
- Silnik inferencji:
  - Dowolny silnik inferencji: vllm, lama.cpp lub ollama,
  - Model open source np. Gemma4 lub klasyfikator np. BERT,
  - Klasyfikacja ryzyka poprzez funkcje scoringowe, która na podstawie warstwy wyjściowej przypisuje materiał do kategorii.
- Logika biznesowa:
  - Agregator wyników, łączący wyniki,1,2,3 w ujednolicone obiekty.

3. Aplikacja Webowa (Dashboard) (React/Vue/Angular):

- Interfejs Analityczny:
  - Wyświetlanie wyników z filtrowaniem po dacie, źródle, poziomie ryzyka i module analitycznym,
  - Wykresy trendów (np. liczba wykrytych incydentów w czasie) za pomocą bibliotek typu Chart.js lub Recharts.
- System Alertowania:
  - Funkcjonalność zmiany statusu rekordu,
  - Implementacja WebSocketów np. Socket.io dla powiadomień o wykryciu wysokiego ryzyka w czasie rzeczywistym.
- Panel Kalibracji:
  - Implementacja parametru progowego (threshold),
  - Zmiana wartości treshold bezpośrednio wpływa na liczbę wyników False Positive vs False Negative (regulacja precyzji i pełności wykrywania).
- UI:
  - Możliwość korzystania z bibliotek do UI open source,
  - Architektura oparta o propsy.

Korzyści z rozwiązania
Stworzony system pozwoli Urzędowi Komisji Nadzoru Finansowego na skokowe zwiększenie efektywności w zakresie:

- Weryfikacji zasad współpracy pomiędzy firmą inwestycyjną a podmiotami reklamującymi (Moduł 1).
- Kontroli transparentności i poprawności informacji o ryzyku (Moduł 2).
- Wykrywania nielegalnego świadczenia usług doradztwa inwestycyjnego na rynku (Moduł 3).
- Wspólnego budowania bezpiecznego rynku finansowego dla wszystkich inwestorów!

---

4. Wymagania formalne

Projekt przesyłany do oceny powinien zawierać:

- szczegółowy opis i tytuł projektu,
- prezentację w formacie PDF,
- opis kodu/readme,
- repozytorium kodu.

Dodatkowo może zawierać:

- krótki film trwający maksymalnie 3 minuty i prezentujący projekt.

---

5. Wymagania techniczne

Projektując rozwiązanie, uczestnicy powinni kierować się następującymi wytycznymi:

- Modułowa architektura lekkich serwisów - separacja Frontendu, Backend-u i silnika AI w niezależne, czytelne moduły, unikając ciężkich struktur monorepo na rzecz przejrzystego podziału.
- Konteneryzacja i szybki start – pełne wykorzystanie Docker i Docker Compose, aby zapewnić uruchomienie całego stosu on-premise jedną komendą, gwarantując powtarzalność środowiska dla kilku osób.
- Stos Open Source i On-Premise – zastosowanie wyłącznie darmowych modeli AI (np. via Ollama, vLLM) oraz otwartych baz danych i bibliotek, całkowicie eliminując zewnętrzne, płatne API.
- Weryfikowalny łańcuch dostaw (Cybersec) – obowiązkowe skanowanie zależności (npm, pip) narzędziami typu Snyk w celu eliminacji podatności oraz ścisła izolacja sekretów w plikach .env. Można to robić na dedykowanych stronach internetowych jeszcze przed pobraniem paczki lub poprzez komendy naprawiające npm audit / npm fix
- Dokumentacja techniczna – README ograniczony do niezbędnego minimum: jasna instrukcja uruchomienia (Quick Start), lista wymagań sprzętowych i opis konfiguracji serwisów, bez zbędnych treści. Bez AI Slop.

---

6. Sposób testowania i/lub walidacji

Jury będzie oceniać projekty podczas prezentacji/ demo projektu. Ocenie podlegać będzie:

- Skuteczność działania algorytmów scrapujących (poprawność pobierania danych ze źródeł)
- Trafność klasyfikacji ryzyka przez moduły AI oraz wpływ zmian parametrów progowych na wyniki w panelu kalibracji.
- Stabilność i płynność działania interfejsu webowego oraz powiadamiań w czasie rzeczywistym.
- Użyte biblioteki i ich bezpieczeństwo.
- Console.log

---

7. Dostępne zasoby

Wszelkie niezbędne materiały takie jak np.: próbki danych zostaną dostarczone uczestnikom bezpośrednio w trakcie trwania hackatonu. Możliwość pobrania danych na pendrive przy stoisku.

---

8. Kryteria oceny

Jurorzy będą oceniać rozwiązania uczestników Konkursu lub Zespoły w skali punktowej od 0 do 10 punktów (maksymalnie 10), biorąc pod uwagę następujące kryteria oceny i wagi dla tych kryteriów:

1) aspekt technologiczny – jak ciekawe jest rozwiązanie, na ile jest gotowe na prezentację wersji demo – waga 30%;
2) innowacyjność wykonania rozwiązania – czy proces wytwarzania oprogramowania opiera się na istniejących standardach, jak opisany jest ustandaryzowany proces tworzenia oprogramowania przy pomocy asystentów AI – waga 20%;
3) architektura rozwiązania – uwzględnia ocenę takich elementów jak: kompozycja modułów systemu, skalowalność, otwartość na rozbudowę – waga 20%
4) na ile rozwiązanie wpasowuje się w kategorię tematyczną zadania konkursowego – w szczególności zgodność ze specyfikacją zadania i preferowanymi technologiami – waga 20%;
5) atrakcyjność rozwiązania – jak bardzo zaskakujące i unikalne w ocenie Jury Konkursu jest rozwiązanie lub praca nad rozwiązaniem problemu (zadania konkursowego) – waga 10%.

---

9. Dodatkowe uwagi / kontekst wdrożeniowy

Najlepsze rozwiązania wyłonione w ramach hackatonu mają szanse na dalszy rozwój, testy pilotażowe lub wdrożenie produkcyjne w strukturach UKNF gdzie mogą zostać wykorzystane w działaniach nadzorczych wspomagających monitorowanie treści w Internecie.

Urząd przewidział nagrody za zajęcie I miejsca UKNF ustanawia nagrodę w Konkursie w wysokości 30 000 zł brutto.

---

10. Kontakt
Mentorzy będą dostępni przy stoisku UKNF.
Dodatkowo do komunikacji zachęcamy do skorzystania z dedykowanego kanału na platformie Discord, Kanał #InvestWatch

---